



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

**Genetic diversity of *Rhus chinensis* and
Eclipta spp. identified by chloroplast
genomes**

BY

INSEO KIM

FEBRUARY, 2018

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

COLLEGE OF AGRICULTURAL AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Genetic diversity of *Rhus chinensis* and *Eclipta* spp. identified by chloroplast genomes

INSEO KIM

DEPARTMENT OF PLANT SCIENCE
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSTRACT

Rhus chinensis is a shrub that belongs to Anacardiaceae family widely distributed in Asia. It is used for medicines and ecological restoration. Also, *Eclipta prostrata* is a medicinal herb belonging to Asteraceae family. It is usually confused with *Eclipta alba* because of their similar morphological features and widely shared habitats. For discrimination of these plants at intra- and inter-species level, we report the complete chloroplast genome of Chinese *R. chinensis*. The assembled chloroplast genome of *R. chinensis* is 149,094bp long consisting of a Large Single copy (97,246bp), a Small Single Copy (18,644bp) and Inverted Repeat (16,602bp) regions. A total of 111 genes in chloroplast genome of *R. chinensis* were revealed including 77 protein coding

genes, 30 tRNA genes, and 4 rRNA genes. A phylogenetic analysis of *R. chinensis* with 11 complete chloroplast genomes in Sapindales order demonstrated the relationship of *R. chinensis* with other species of Sapindales. A comparative analysis of *R. chinensis* revealed 170 SNPs and 85 InDels at intra-species level. Based on the sequence variations between Chinese and Korean collections, we developed 3 molecular markers that could identify the genetic diversity of *R. chinensis* at intra-species level. Likewise, the complete chloroplast genome of *E. alba* were generated in this study. The assembled chloroplast genome of *E. alba* is 151,733bp long composed of the same structure as *R. chinensis*. A Large Single Copy, a Small Single Copy, and Inverted Repeat regions are 83,300bp, 18,283, and 25,075bp long, respectively. Gene annotation revealed 80 protein coding genes, 30 tRNA genes, and 4 rRNA genes. A phylogenetic analysis of *E. alba* with 12 complete chloroplast genome sequences in Asteraceae family clarified the relationship of *E. alba* with other species of Asteraceae family. A comparative analysis exhibited 58 InDels and 29 SNPs at inter-species level between *E. prostrata* and *E. prostrata*. A total of 6 molecular markers were developed with these polymorphic sites that could successfully distinguish *E. prostrata* and *E. alba*. The chloroplast genome sequences of medicinal plants generated in this study and the molecular markers developed for discrimination and authentication will not only sort out the right plant individual or species at intra- or inter-species level but also provide valuable information for further genetic researches.

Key words: *Rhus chinensis*; *Eclipta alba*; The complete chloroplast genome; Phylogenetic analysis; molecular marker; intra- or inter-species; demonstration

Student number: 2016-21352

CONTENTS

| | |
|---|------------|
| ABSTRACT | I |
| CONTENTS..... | III |
| LIST OF TABLES | V |
| LIST OF FIGURES | VI |
| INTRODUCTION..... | 1 |
| I. The Complete Chloroplast Genome Sequence and Intra-Species Diversity of <i>Rhus chinensis</i> | 3 |
| 1. Introduction | 3 |
| 2. Materials and Methods | 4 |
| 2.1. Plant materials and genome sequencing | 4 |
| 2.2. Chloroplast genome assembly | 4 |
| 2.3. Phylogenetic analysis | 6 |
| 2.4. Comparison of intra-species level and development of molecular marker | 6 |
| 3. Results | 8 |
| 3.1. Complete chloroplast genome sequences and sequence variations of <i>R. chinensis</i> | 8 |
| 3.2. Phylogenetic analysis of <i>R. chinensis</i> based on the chloroplast genomes..... | 13 |
| 3.3 Molecular markers for classification of Chinese and Korean <i>R. chinensis</i> | 15 |
| 4. Discussion | 20 |
| 4.1. Intra-species diversity of <i>R. chinensis</i> | 20 |
| 4.2. Discrimination of <i>R. chinensis</i> collections with three molecular markers | 20 |

| | |
|---|-----------|
| II. Discrimination and Authentication of <i>E. prostrata</i> and <i>E. alba</i> Based on the Complete Chloroplast Genomes | 22 |
| 1. Introduction | 22 |
| 2. Materials and Methods | 24 |
| 2.1. Plant materials and genome sequencing | 24 |
| 2.2. Chloroplast genome assembly | 24 |
| 2.3. Phylogenetic analysis | 26 |
| 2.4. Comparison of inter-species level and development of molecular marker | 26 |
| 3. Results | 28 |
| 3.1. Complete chloroplast genome sequence of <i>E. alba</i> and sequence variations between <i>E. prostrata</i> and <i>E. alba</i> | 28 |
| 3.2. Phylogenetic analysis of <i>E. alba</i> based on the chloroplast genomes | 35 |
| 3.3 Molecular markers for discrimination of <i>E. prostrata</i> and <i>E. alba</i> & Application of InDel markers to 5 random specimens | 37 |
| 4. Discussion | 47 |
| 4.1. Discrimination of <i>E. prostrata</i> and <i>E. alba</i> with six molecular markers | 47 |
| 4.2. Authentication of <i>Eclipta</i> products with three InDel-based markers | 47 |
| REFERENCES | 49 |
| ABSTRACT IN KOREAN | 55 |

LIST OF TABLES

| | | |
|-----------------|--|----|
| Table 1. | Eight <i>R. chinensis</i> collections used in this study | 5 |
| Table 2. | Summary of NGS data and chloroplast genome sequences of <i>R. chinensis</i> collections | 9 |
| Table 3. | Annotated genes in chloroplast genome of <i>R. chinensis</i> | 11 |
| Table 4. | Information of molecular markers developed in this study for discrimination of <i>R. chinensis</i> | 16 |
| Table 5. | <i>E. prostrata</i> and <i>E. alba</i> used in this study | 25 |
| Table 6. | Summary of NGS data and chloroplast genome sequences of <i>E. prostrata</i> and <i>E. alba</i> | 30 |
| Table 7. | Annotated genes in chloroplast genome of <i>E. alba</i> | 32 |
| Table 8. | Information of molecular markers developed in this study for classification between <i>E. prostrata</i> and <i>E. alba</i> | 38 |

LIST OF FIGURES

| | | |
|-------------------|---|----|
| Figure 1. | Chloroplast genome map of <i>R. chinensis</i> | 10 |
| Figure 2. | Comparison of chloroplast genomes of two <i>R. chinensis</i> | 12 |
| Figure 3. | Phylogenetic analysis of <i>R. chinensis</i> | 14 |
| Figure 4. | InDel markers for classification of <i>R. chinensis</i> at intra-species level..... | 17 |
| Figure 5. | The SNP marker for discrimination of <i>R. chinensis</i> at intra-species level..... | 19 |
| Figure 6. | Chloroplast genome map of <i>E. alba</i> | 31 |
| Figure 7. | Comparison of chloroplast genome sequences of <i>E. prostrata</i> and <i>E. alba</i> | 33 |
| Figure 8. | Phylogenetic analysis of <i>E. alba</i> | 36 |
| Figure 9. | InDel markers for discrimination between <i>E. prostrata</i> and <i>E. alba</i> at inter-species level..... | 39 |
| Figure 10. | SNP markers for classification between <i>E. prostrata</i> and <i>E. alba</i> at inter-species level..... | 41 |
| Figure 11. | Application of InDel markers to random specimens of <i>Eclipta</i> species | 45 |

INTRODUCTION

Molecular markers have been widely used as tools for detection of genome differences in scientific field. Among the rest, the molecular markers have cooperated in prospering botany. For example, researchers generated genetic map using the markers in *Phaseolus vulgaris* (Grisi *et al.* 2007) and *Populus deltoides* (Zhou *et al.* 2015), which can provide information to genetic breeding. Also, expression of genome in plants can be identified by molecular markers. Transcriptome assembly and gene expression profiling had been conducted in *Onobrychis viciifolia* (Mora-Ortiz *et al.* 2016) and *Elaeis guineensis* (Ling *et al.* 2016). Meanwhile, molecular markers are useful for authentication of plant species. Many previous studies had been conducted for discrimination with the markers in *Cynanchum wilfordii*, *Cynanchum auriculatum* and *Polygonum multiflorum* (Kim *et al.* 2015c; Jang *et al.* 2015).

For exact identification of plant species, DNA barcoding have been suggested for land plants (Hollingsworth *et al.* 2009). In previous research, rbcL+matK were selected as standard regions for effective classification among plant species (Hollingsworth *et al.* 2009). However, DNA barcode markers cannot fully cover the plant phylogeny because they are limited to amplify certain regions. Dong *et al.* (2012) presented supplementations of disadvantages in DNA barcode markers. Researchers could use the chloroplast genomes to analyse the genetic diversity and differentiation of plant species at low taxonomic levels in depth more than universal barcode markers.

Chloroplast DNA carries a feature of containing polymorphisms at intra- and inter-species level that makes itself a good candidate for classification among populations and species. Chloroplast genomes are multifunctional organelles carrying their own genetic materials. In most of higher plants,

chloroplasts have circular structure and their sizes range from 120 to 160kb. The circular formats of chloroplast genomes possess a quadripartite structure composed of two Inverted Repeat (IR) regions, Large Single Copy (LSC) region, and Small Single Copy (SSC) region (Jansen *et al.* 2005).

Chloroplast genomes are typically uni-parentally inherited (Cheng *et al.* 2005; Dong *et al.* 2012) and rarely apt to occur recombination and mutation. Due to these characteristics, chloroplast genomes have been widely used for evolutionary researches at intra- or inter-species level. For example, discrimination at intra-species level had been conducted with using chloroplast DNA in plant species such as *Panax ginseng*, *Pedicularis chamissonis*, *Primula cuneifolia*, *Tellima grandifolia*, and *Tiarella trifoliata* (Soltis *et al.* 1991; Soltis *et al.* 1992; Fujii *et al.* 1995; Sewell *et al.* 1996; Fujii *et al.* 1997; Kim *et al.* 2015a; Joh *et al.* 2017). Meanwhile, phylogenetic researches had been conducted at inter-species level based on chloroplast genomes in *Scalesia affinis*, *Abies* species, and *Citrus* (Parducci *et al.* 2000; Nielsen *et al.* 2004; Cheng *et al.* 2005). Also, *Rhus* species had been discriminated with using chloroplast genomes and 45S nrDNA genome (Miller *et al.* 2001; Lee *et al.* 2004; Yi *et al.* 2004; Yi *et al.* 2007). Likewise, development of chloroplast DNA markers might also allow *Rhus chinensis* and *Eclipta prostrata* to be distinguished among their populations at intra-species level and *Eclipta alba* at inter-species level, respectively.

In this study, the complete chloroplast genomes of two *R. chinensis* genotypes from China and Korea and *E. alba* were assembled. Then, DNA markers were developed based on the polymorphic regions of chloroplast genomes that could successfully be applied for authentication and classification of *R. chinensis* collections & *E. prostrata* and *E. alba*.

I . The Complete Chloroplast Genome Sequence and Intra-Species Diversity of *Rhus chinensis*

1. Introduction

Rhus chinensis is a deciduous shrub belonging to the family Anacardiaceae and distributed widely in Asia including India, Vietnam, China, Korea, and Japan (Min and Barfod 2008). It contains various pharmacologically active constituents and insect-induced galls, which have been used for medicinal purposes (Min and Barfod 2008; Djakpo and Yao 2010). In addition, *R. chinensis* is used as revegetation plant for ecological restoration owing to its cold tolerance and easy multiplication by both seed and clonal propagation (Nam *et al.* 2004; Lim and Oh 2015). The diverse utilization of this shrub led to international seed trading which in turn resulted in contamination and destruction of endemic population of *R. chinensis*. Although the generic and infrageneric delimitation of *Rhus* species remains controversial, very limited genetic resources are available for *R. chinensis* (Young 1978; Miller *et al.* 2001; Yi *et al.* 2004; Ma *et al.* 2013; Lee *et al.* 2016). In this study, we have characterized the complete chloroplast genome of Chinese *R. chinensis* and compared with Korean *R. chinensis* chloroplast genome for estimating genetic diversity and developing marker for authentication.

In this study, we assembled the complete chloroplast genome of two *R. chinensis* genotypes from China and Korea. In addition, we have conducted a comparative phylogenomic analysis of Chinese and Korean *R. chinensis* along with 11 species belonged in the Sapindales order. We also developed polymorphic DNA markers derived from chloroplast genomes to practically apply for authentication and genetic diversity among *R. chinensis* collections.

2. Materials and Methods

2.1. Plant materials and genome sequencing

Seeds of wild Chinese *R. chinensis* were obtained from Shandong and Henan provinces in China. We also collected seven *R. chinensis* samples in Korea (Table 1): two samples from demilitarized zone (DMZ) between South and North Korea and five samples from other locations in South Korea. We sequenced plants collected from Yang-gu province in Korea and Shandong and Henan provinces in China as representatives of the collections of Korea and China. The genomic DNAs were extracted from the leaf tissues or seeds of collected plants using a modified cetyltrimethylammonium bromide (CTAB) method (Allen *et al.* 2006) and quantified using Nanodrop ND-1000 (Nanodrop Technologies, Inc., Wilmington, DE, USA). Paired-end (PE) sequencing was conducted using Illumina MiSeq platform by LabGenomics (www.labgenomics.co.kr, Seongnam, Korea).

2.2. Chloroplast genome assembly

Raw PE reads were trimmed and de novo assembled using the method as described by Kim *et al.* (2015a, 2015b). From initial assembly, contigs representing chloroplast genome sequences were extracted, ordered, and merged to generate a single contig sequence using the reference chloroplast genome *Acer buergerianum* ssp. *ningpoense* (KF753631, Yang *et al.* 2014). The assembled sequence was manually corrected and gap-filled by a series of PE read mapping. The assembled chloroplast genome was annotated using GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq-app.html>) and manually curated using the Artemis annotation tool (Rutherford *et al.* 2000).

Table 1. Eight *R. chinensis* collections used in this study

| No. | Species | Countries | Collected locations | Collected year |
|-----|---------------------|--------------------|--|----------------|
| 1 | <i>R. chinensis</i> | China ^a | Shandong and Henan provinces | 2014 |
| 2 | | Korea ^b | Yang-gu, Gangwon-do | 2014 |
| 3 | | Korea ^b | Hwacheon, Gangwon-do | 2014 |
| 4 | | Korea ^b | Mt.Jiri, Sancheong, Kyungdsangnam-do | 2014 |
| 5 | | Korea ^b | Han-taek Botanical Garden, Yong-in, Kyungki-do | 2015 |
| 6 | | Korea ^b | Mt. Kariwang, Jeong-seon, Gangwon-do | 2015 |
| 7 | | Korea ^b | Gwangreung Botanical Garden, Korea National Arboretum, Po-cheon, Gyeonggi-do | 2015 |
| 8 | | Korea ^b | Gwangreung Botanical Garden, Korea National Arboretum, Po-cheon, Gyeonggi-do | 2015 |

^a Seeds

^b Leaves

2.3. Phylogenetic analysis

Phylogenetic analysis was carried out using multiple sequence alignments of 13 complete chloroplast genome sequences (*Acer burgerianum* ssp. *ningpoense*, KF753631; *Acer davidii*, NC_030331; *Acer miaotaiense*, NC_030343; *Azadirachta indica*, NC_023792; *Boswellia sacra*, KT934315; *Citrus aurantiifolia*, KJ865401; *Citrus sinensis*, DQ864733; *Dipteronia sinensis*, NC_029338; *R. chinensis* (China), MF351625; *R. chinensis* (Korea), NC_033535; *Sapindus mukorossi*, KM454982; *Spondias bahiensis*, KU756561; *Spondias tuberosa*, KU756562) that belong to the Sapindales order. Phylogenetic tree was constructed using MEGA6.0 (Tamura *et al.* 2013) with the parameters of neighbor-joining method and 1000 bootstrap replicates.

2.4. Comparison of intra-species level and development of molecular marker

Sequence variations were identified by mVISTA program (<http://genome.lbl.gov/vista/mvista/submit.shtml>) and the polymorphic sites were arranged by MAFFT program (<http://mafft.cbrc.jp/alignment/software>). The two types of molecular markers, InDel and SNP, were developed based on the polymorphic sites in chloroplast genomes of two *R. chinensis*. The primers were designed using Primer-blast tool in NCBI (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). PCR reactions were performed in 25 μ L final volume which is composed of 20 ng of template DNA, 1x Taq buffer, 2.5 mM dNTP, 10 pmol of each primer, and 2 unit/ μ L Taq DNA polymerase (Vivagen, Korea). PCR conditions were as follows: 5 minutes at 95°C, 35 cycles of 30 seconds at 95°C, 30 seconds at 56°C, and 30 seconds at 72°C, and 5 minutes at 72°C as final extension. PCR amplicons were inspected using 3% agarose gel including Inclone™ Safe Gel stain. Then they were visualized under UV trans-illuminator and a gel documentation system.

PCR reaction for high resolution melting (HRM) analysis against SNP target was performed in a 20 μ L final volume which consists of 20 ng of template DNA,

1x Taq buffer, 2.5 mM dNTP, 10 pmol of each primer, 2 unit/ μ L Taq DNA polymerase (Vivagen, Korea), and fluorescent dye SYTO 9 (Roche Diagnostics). PCR conditions were as follows: 5 minutes at 95°C, 45 cycles of 30 seconds at 95°C, 30 seconds at 56°C, and 30 seconds at 72°C, and 5 minutes at 72°C as final extension. Then, HRM analysis was conducted using LightCycler 480 (Roche Applied Science). HRM conditions were as follows: 1 minute at 95°C, 1 minute at 40°C, 5 seconds at 70°C, and the temperature increased up to 90°C, then decreased to 40°C as fluorescence acquisition.

3. Results

3.1. Complete chloroplast genome sequences and sequence variations of *R. chinensis*

Paired-end (PE) sequencing was conducted with Illumina Miseq platform in Chinese *R. chinensis* to produce 4.1Gbp raw reads (Table 2). The chloroplast genome of Chinese *R. chinensis* was assembled using *de novo* assembly method in previous study (Kim *et al.*, 2015a; Kim *et al.*, 2015b).

The assembled chloroplast genome sequence of Chinese *R. chinensis* is 149,094bp long and retains circular quadripartite structure consisting of Large Single Copy (LSC), Small Single Copy (SSC), and two Inverted Repeat (IR) regions. The lengths of each part are 97,246bp, 18,644bp and 16,602bp long, respectively (Fig. 1, Table 2). A total of 111 genes were annotated in the chloroplast genome of Chinese *R. chinensis* including 77 protein coding genes, 30 tRNA genes, and 4 rRNA genes (Table 3). Also, GC content of the chloroplast genome of Chinese *R. chinensis* was 37.86% (data not shown).

Comparison to chloroplast sequences of Chinese *R. chinensis* and Korean *R. chinensis* revealed 83bp differences. In LSC region, Chinese *R. chinensis* is 364bp longer than Korean *R. chinensis*. However, Korean *R. chinensis* is 139bp and 3bp longer in IR and SSC regions, respectively (Table 2).

Intra-species polymorphisms of chloroplast genomes between two types of *R. chinensis* were identified using mVISTA program (Fig. 2). Intra-species polymorphic sites are positioned in intergenic regions more than genic regions. A total of 170 SNPs and 85 InDels between two types of *R. chinensis* were identified. The analysis of sequence variations revealed 99 SNPs and 78 InDels were located in intergenic regions. Also, 71 SNPs and 7 InDels were identified from genic regions. These polymorphic sites were used for phylogenetic analysis of *R. chinensis* and development of molecular markers (Fig. 2).

Table 2. Summary of NGS data and chloroplast genome sequences of *R. chinensis* collections.

| Collected species | Raw reads (bp) | Cp coverage (x) | Cp length (bp) | LSC length (bp) | SSC length (bp) | IR length (bp) |
|---|-------------------|--------------------|-------------------|--------------------|--------------------|-------------------|
| <i>R. chinensis</i> (China) | 4,199,252,923 | 162 | 149,094 | 97,246 | 16,602 | 18,644 |
| <i>R. chinensis</i> (Korea) ^a | 6,251,344,649 | 96 | 149,011 | 96,882 | 16,741 | 18,647 |

^a conducted by Lee *et al.* (2016)

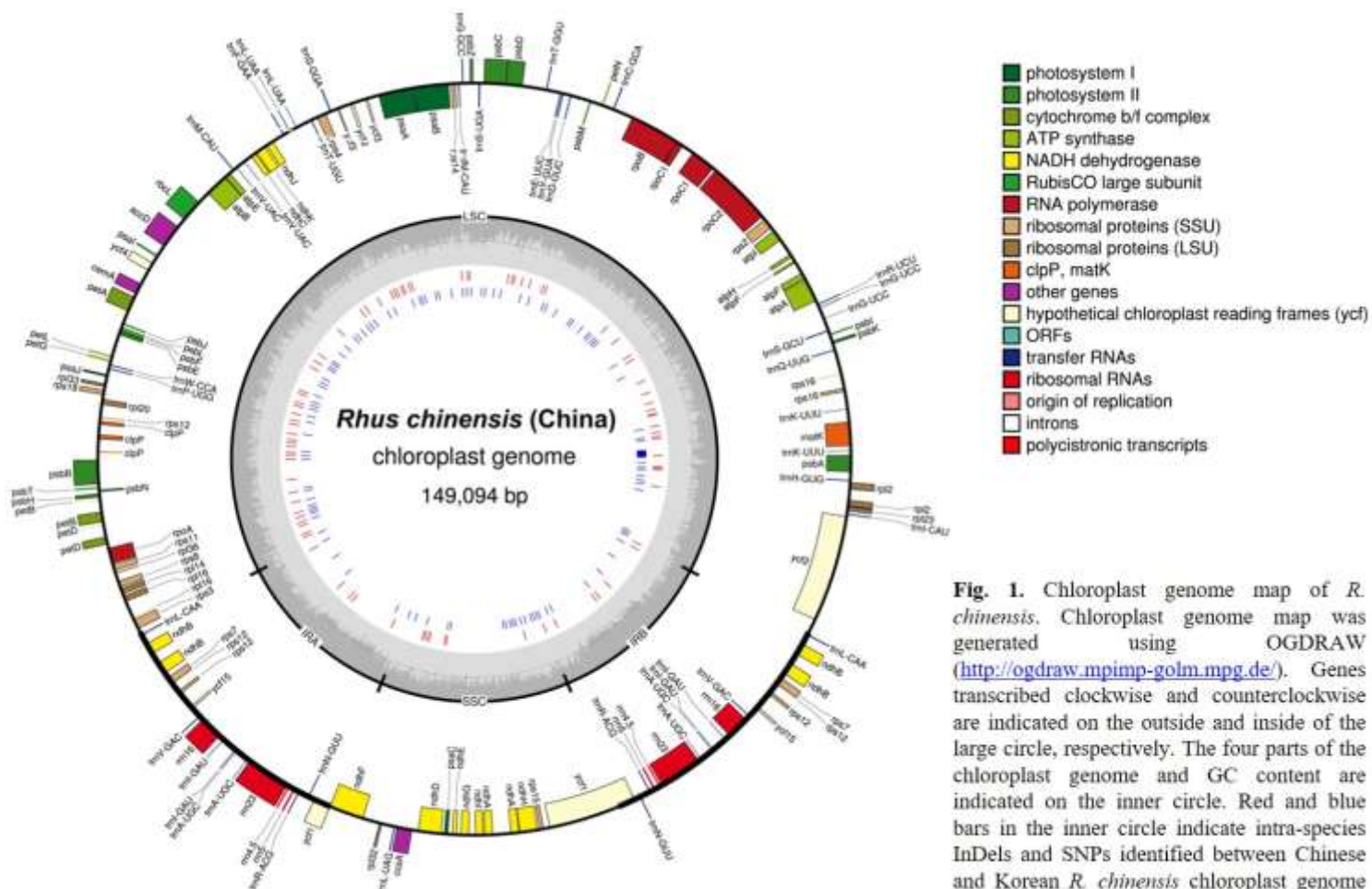


Fig. 1. Chloroplast genome map of *R. chinensis*. Chloroplast genome map was generated using OGDRAW (<http://ogdraw.mpimp-golm.mpg.de/>). Genes transcribed clockwise and counterclockwise are indicated on the outside and inside of the large circle, respectively. The four parts of the chloroplast genome and GC content are indicated on the inner circle. Red and blue bars in the inner circle indicate intra-species InDels and SNPs identified between Chinese and Korean *R. chinensis* chloroplast genome sequences.

Table 3. Annotated genes in chloroplast genome of *R. chinensis*

| Gene types | Gene list |
|---|--|
| Photosystem I | <i>psaA, B, C, I, J</i> |
| Photosystem II | <i>psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z</i> |
| Cytochrome b/f complex | <i>petA, B, D, G, N, L</i> |
| ATP synthase | <i>atp A, B, E, F, H, I</i> |
| NADH dehydrogenase | <i>ndhA, B, C, D, E, F, G, H, I, J, K</i> |
| RubisCO large subunit | <i>rbcL</i> |
| RNA polymerase | <i>rpoA, B, C1, C2</i> |
| Ribosomal proteins (SSU) | <i>rps2, 3, 4, 7, 8, 11, 12, 14, 15, 16, 18</i> |
| Ribosomal proteins (LSU) | <i>rpl2, 14, 16, 20, 23, 32, 33, 36</i> |
| clpP, matK | <i>clpP, matK</i> |
| Other genes | <i>ccsA, cemA, accD</i> |
| Hypothetical chloroplast reading frames | <i>ycf1, 2, 3, 4, 15</i> |
| Transfer RNAs | <i>trnFM-CAU, A-UGC, C-GCA, D-GUC, E-UUC, F-GAA, G-GCC,UCC, H-GUC, I-CAU,GAU, K-UUU, L-CAA,CAU,UAG,UAA, M-CAU, N-GUU, P-UGG, Q-UUG, R-UCU,ACG,S-UGA,GCU,GGA, T-GGU, V-GAC,UAC,W-CCA, Y-GUA</i> |
| Ribosomal RNAs | <i>rrn4.5, 5, 16, 23</i> |

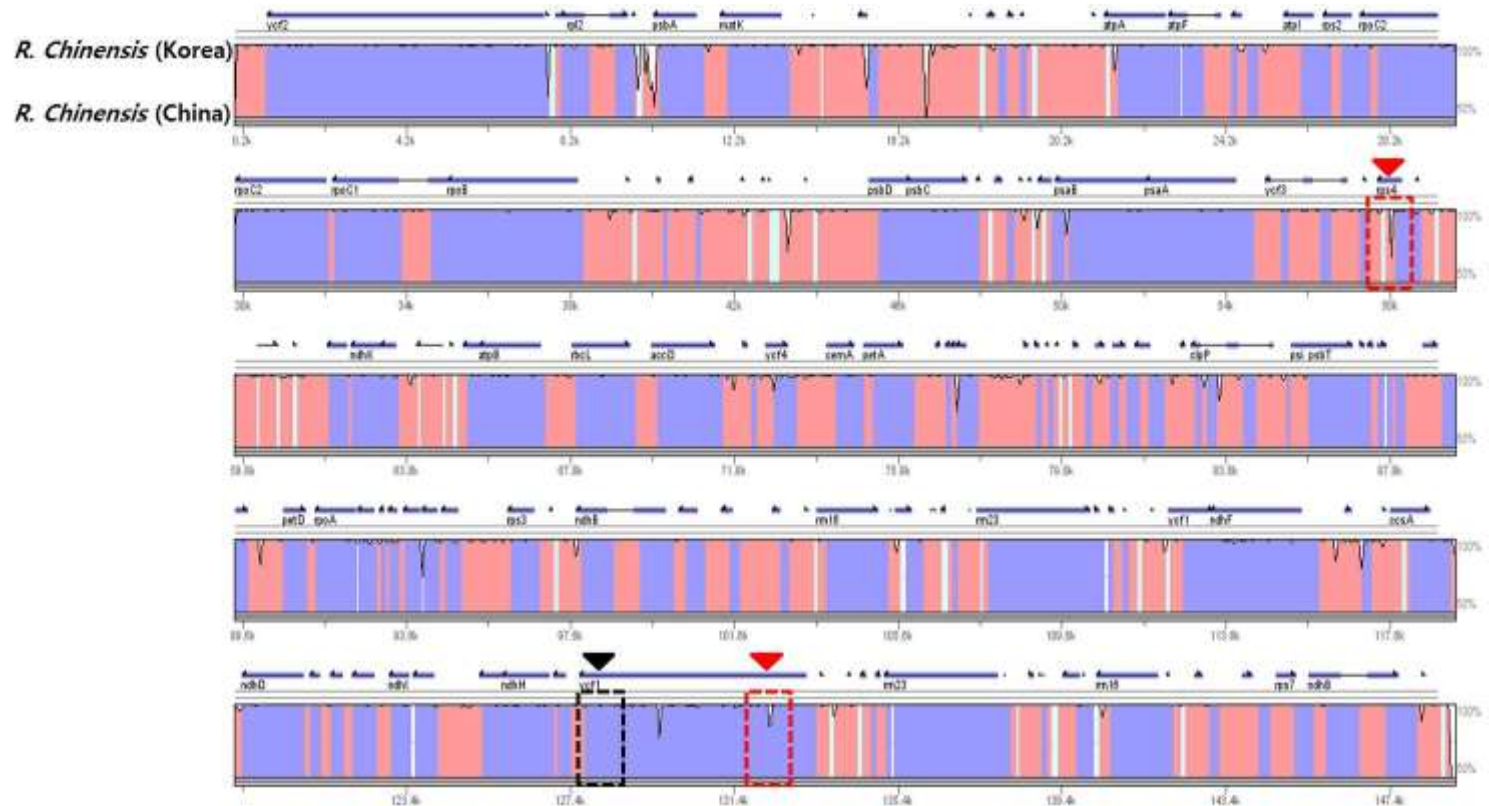


Fig. 2. Comparison of chloroplast genomes of two *R. chinensis*. Genic regions were revealed by GeSeq and Artemis annotation tool. Comparative map was generated with Korean *R. Chinensis* as a reference using mVISTA. The polymorphic regions used for molecular marker development are represented by arrowheads. Red arrowheads represent the regions used for InDel marker development. Black arrowhead shows the site used for SNP marker development.

3.2. Phylogenetic analysis of *R. chinensis* based on the chloroplast genomes

Phylogenetic analysis was performed with two types of the complete chloroplast genomes of *R. chinensis* and 11 complete chloroplast sequences of Sapindales order. As expected, the phylogenetic tree demonstrated that *R. chinensis* was located in the same group of Spondias species of Anacardiaceae family (Fig. 3). Meanwhile, the group including *R. chinensis* revealed a close sister relationship with the species of Burseraceae and Sapindaceae family in Sapindales order. This analysis is in accordance with previous study of Korean *R. chinensis* (Lee *et al.*, 2016).

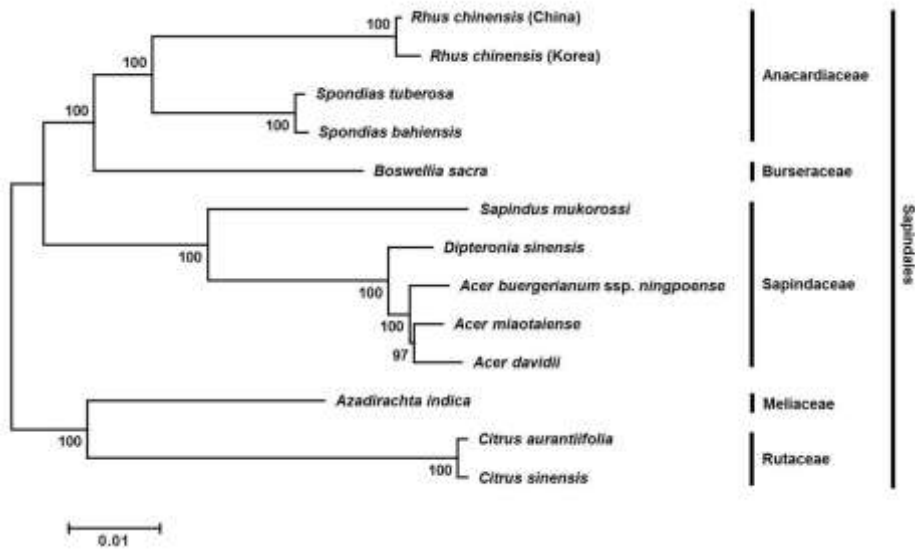


Fig. 3. Phylogenetic analysis of *R. chinensis*. The tree was generated with complete chloroplast genome sequences of *R. chinensis* and species belonging to the Sapindales order by multiple alignment using MAFFT (<http://mafft.cbrc.jp/alignment/server/index.html>) and a neighbor-joining (NJ) analysis using MEGA 6.0 (Tamura *et al.*, 2013). Numbers in the nodes are bootstrap support values (>50%) from 1000 replicates. Chloroplast genome sequences used for this tree are: *Acer burgerianum* ssp. *ningpoense*, KF753631; *Acer davidii*, NC_030331; *Acer miaotaiense*, NC_030343; *Azadirachta indica*, NC_023792; *Boswellia sacra*, KT934315; *Citrus aurantiifolia*, KJ865401; *Citrus sinensis*, DQ864733; *Dipteronia sinensis*, NC_029338; *R. chinensis* (China), MF351625; *R. chinensis* (Korea), NC_033535; *Sapindus mukorossi*, KM454982; *Spondias bahiensis*, KU756561; *Spondias tuberosa*, KU75656

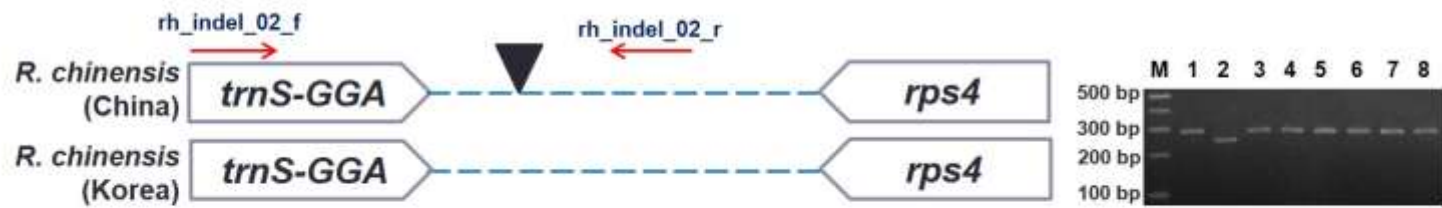
3.3. Molecular markers for classification of Chinese and Korean *R. chinensis*

The molecular markers developed in this study for discrimination of *R. chinensis* at intra-species level are described in Table 4. One of the InDel markers, rh_indel_02, successfully amplified the PCR products, and the amplicons showed 39-bp differences in *trnS-GGA-rps4* region between Chinese and Korean (Yang-gu) *R. chinensis* as we expected (Fig. 4(A)). Likewise, amplicons produced by rh_indel_03 marker demonstrated 25-bp InDel polymorphism in *ycf1* region between two types of *R. chinensis* as expected (Fig. 4(B)). Similarly, when rh_hrm_11, the SNP marker, applied to the samples of *R. chinensis*, HRM analysis showed different melting patterns of Chinese *R. chinensis* not only from Korean *R. chinensis* from Yang-gu province but other areas. Also, the same sizes of the amplicons produced from rh_hrm_11 validate there are no any other polymorphisms but SNPs in target sites of the marker (Fig. 5).

Table 4. Information of molecular markers developed in this study for discrimination of *R. chinensis*

| Type | Marker ID | Location | | Primer Sequence(5'-3') | Product Size (bp) |
|-------|-------------|----------------------|---|-------------------------|-------------------|
| InDel | rh_indel_02 | <i>trnS-GGA-rps4</i> | F | AGTGGTTCAAGGCGTAGCAT | 288 |
| | | | R | ATTTGATCCGGCGATTGGA | |
| | rh_indel_03 | <i>ycfI</i> | F | TGATTCGCTCGATTTCGCCA | 221 |
| | | | R | TCTGTCCTTCAATATCACGGAAC | |
| SNP | rh_hrm_11 | <i>ycfI</i> | F | CATGTGTGCATCTCTGGGTT | 196 |
| | | | R | CTTCCTTTGGTCCAATTCTCGAT | |

(A)



(B)

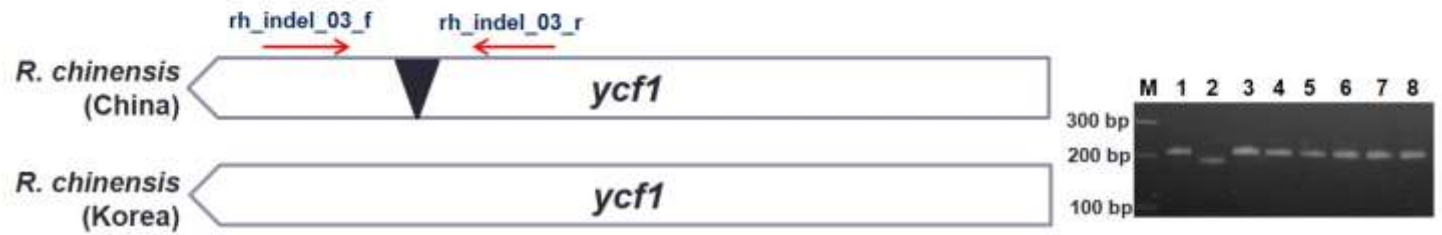


Fig. 4. InDel markers for classification of *R. chinensis* at intra-species level. (A) The marker, rh_indel_02 showed the amplicons that include 39-bp InDel (GGGGAGAAAAATCTTCCTACTGATCTATGATACATGAAT) in *trnS-GGA-rps4* region. (B) The marker, rh_indel_03 showed the amplicons that contain 25-bp InDel (AAAATTCCAAAAAGGTTTATGTCTT) in *ycf1* region. *R. chinensis* collections; 1: China (Shandong and Henan), 2: Korea (Yang-gu), 3: Korea (Hwacheon), 4: Korea (Mt. Jiri), 5: Korea (Han-taek Botanical Garden), 6: Korea (Mt. Kariwang), 7, 8: Korea (Gwangreung Botanical Garden), M: 100-bp DNA ladder.

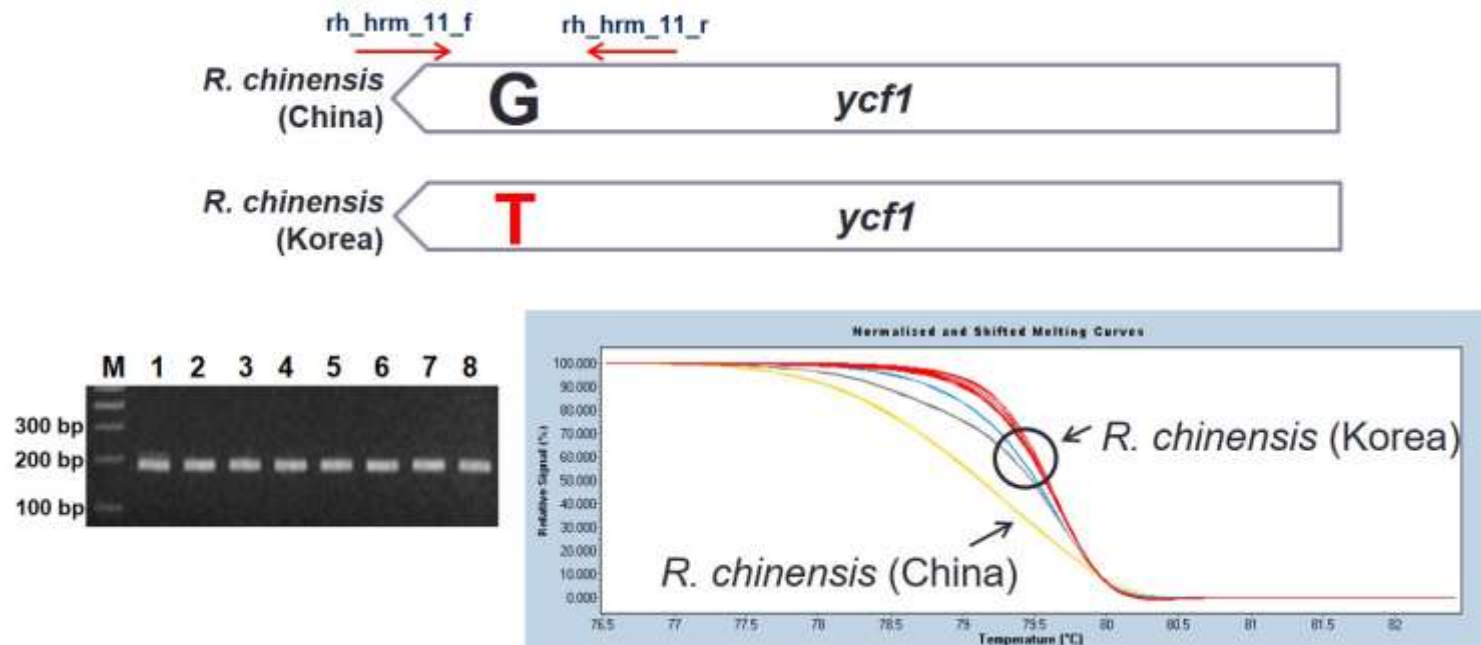


Fig. 5. The SNP marker for discrimination of *R. chinensis* at intra-species level. The marker, rh_hrm_11 showed different melting patterns between Korean *R. chinensis* and Chinese *R. chinensis* in *ycf1* region. *R. chinensis* collections; 1: China (Shandong and Henan), 2: Korea (Yang-gu), 3: Korea (Hwacheon), 4: Korea (Mt. Jiri), 5: Korea (Han-taek Botanical Garden), 6: Korea (Mt. Kariwang), 7, 8: Korea (Gwangreung Botanical Garden), M: 100-bp DNA ladder.

4. Discussion

4.1. Intra-species diversity of *R. chinensis*

R. chinensis is known as a valuable medicinal plants with remedial components (Min and Barfod 2008; Djakpo and Yao 2010). In addition, *R. chinensis* is used for ecological restoration (Nam *et al.* 2004; Lim and Oh 2015). Due to its wide utility, it is necessary to explore this plant for genetic diversity and marker development to prevent false trading (Khairallh and Salama 2009). Here, we generated the complete chloroplast genome of two *R. chinensis* plants collected from Korea and China. Comparative analysis of chloroplast genome of *R. chinensis* revealed 255 intra- species polymorphic sites, 170 SNPs and 85 InDels. The abundant genetic diversity would be found in the wild collections of *R. chinensis* and could be applied to phylogenetic analysis and development of molecular markers for verifying genetic diversity of *R. chinensis*.

4.2. Discrimination of *R. chinensis* collections with three molecular markers

Two InDel markers were unique for one Korean collection of *R. chinensis* (Yang-gu) from the others (Fig. 4(A), (B)). The Yang-gu collection is derived from the DMZ region between South Korea and North Korea. The eco-system was well maintained in DMZ regions for longer than 60 recent years with less artificial interruption. Two InDel markers revealed unique genotype of Yang-gu collection among the eight collections, suggesting the Yang-gu collection are relatively isolated in the DMZ region from others. However, it is required to expand our study by evaluating more population for solid conclusion. Meanwhile, the SNP markers could efficiently divide the melting patterns for Chinese *R. chinensis* from Korean *R. chinensis* through HRM analysis (Fig. 5).

Chloroplast genomes are valuable genomic resources for practical

application of DNA markers because of the uni-parentally inherited features and conserved structure of the chloroplast genomes (Cheng *et al.* 2005; Dong *et al.* 2012, Kim *et al.* 2015a, 2015b, 2015c; Joh *et al.* 2017). Here, we report three DNA markers which can be applied for genetic diversity and practical application for authentication of the plant collections of *Rhus* species.

II . Discrimination and Authentication of *E. prostrata* and *E. alba* Based on the Complete Chloroplast Genomes

1. Introduction

E. prostrata is an annual herbaceous plant belonging to the family Asteraceae which includes 1,620 genera and 23,600 species (Mithun *et al.* 2011). It is distributed in tropical and subtropical regions including South America, Asia, Africa, and south of the central region in Korea (Chokotia *et al.* 2013). It has tap roots and many brown branches. Also, it has two types of flowers, ray florets and disc florets, with involucre of bracts that arrange in two lines (Neeraja and Margaret 2012). *E. alba* is also included in the family Asteraceae and has very similar shapes with *E. prostrata* except for smoother margins of leaf blades. *E. prostrata* has been used as a medicinal plant owing to compounds like Eclalbasaponin I and wedelolactone found in the plant which have antitumor and cirrhosis and hepatitis curing effects (Dalal *et al.* 2010; Liu *et al.* 2012). In addition, it has been used as resources for chinese tonics to cure loose teeth, tinnitus, hemoptysis, hematuria and uterine bleeding (Chinese Pharmacopoeia Commission 2010). Similarly, *E. alba* has been used as a medicinal plant owing to its flavonoids that could induce anagen, which helps blackening hairs (Datta *et al.* 2009). Due to similar morphological traits and widely shared habitats (Baskaran and Jayabalan 2005; Dhaka and Kothari 2005), *E. alba* is often mixed with *E. prostrata* (Muruganantham *et al.* 2009), which cause confusion in the academic world as well as the markets (Neeraja and Margaret 2012). Therefore, it is necessary to develop a high-throughput system to discriminate them for efficient genetic research and sustain its market value.

Previously, we reported the chloroplast genome for *E. prostrata* (Park *et al.* 2016). In this study, we assembled the complete chloroplast genome of *E. alba*. In addition, we have conducted a comparative phylogenetic analysis of *E. prostrata* and *E. alba* along with 10 species that belong to the Asteraceae family. In addition, we show genome-level diversity between the closely related *Eclipta* species and have developed polymorphic markers that could efficiently distinguish *E. prostrata* and *E. alba* and thus can practically be applied for authentication of those species for adequate use of these medicinal herbs.

2. Materials and Methods

2.1. Plant materials and genome sequencing

E. prostrata and *E. alba* plants used in this study were collected from HanTaek Botanical Garden (Yongin, Korea, www.hantaek.co.kr) and Jeju Island, respectively (Table 5). Specimens of dried *Eclipta* plant tissues were provided by Ministry of Food and Drug Safety. Genomic DNAs of *E. prostrata* and *E. alba* were extracted from leaf tissues by a modified cetyltrimethylammonium bromide (CTAB) method (Allen *et al.* 2006), and the DNA of the specimens were extracted from each of the whole, stem, and leaf tissues by Genomic Plus DNA Prep kit (Inclone, Korea). The whole tissues of specimens were composed of stems, leaves, and flowers. The quality and quantity of extracted DNA were checked using Nanodrop ND-1000 (Thermo scientific, Wilmington, USA). Paired-end sequencing (PE) was conducted using Illumina MiSeq platform by LabGenomics (www.labgenomics.co.kr, Seongnam, Korea).

2.2. Chloroplast genome assembly

Raw paired-end (PE) reads were trimmed and de novo assembled using the method as described by Kim *et al.* (2015a, 2015b). From initial assembly, contigs representing chloroplast genome were retrieved, ordered, and combined into a single sequence by comparing chloroplast genome sequence of *Centaurea diffusa* (KJ690264). The assembled sequence was manually corrected and gap-filled by a series of PE read mapping. The assembled chloroplast genome was annotated using DOGMA (<http://dogma.ccbb.utexas.edu>) and manually corrected with BLAST search.

Table 5. *E. prostrata* and *E. alba* used in this study

| No. | Species | Countries | Collected locations | Collected year |
|-----|---------------------|-----------|--|----------------|
| 1 | <i>E. prostrata</i> | Korea | Han-taek Botanical Garden, Yong-in, Kyungki-do | 2015 |
| 2 | <i>E. alba</i> | Korea | Jeju Island | 2015 |

2.3. Phylogenetic analysis

Phylogenetic analysis was conducted using multiple sequence alignments of 29 protein coding gene sequences (*psaI/J*, *psbB/C/D/H/I/K/T*, *petA/B/D/G/N/L*, *atpA/F/H/I*, *ndhB*, *rpoB/C2*, *rps2/7/18*, *rpl2*, *accD*, *ycf2/4*) in 12 chloroplast genomes (*Chrysanthemum* × *morifolium*, JQ362483; *Chrysanthemum indicum*, JN867589; *Artemisia montana*, KF887960; *Artemisia frigida*, JX293720; *Aster spathulifolius*, KF279514; *Leontopodium leirolepis*, KM267636; *Helianthus annuus*, DQ383815; *Eclipta prostrata*, KU361242; *Eclipta alba*, MF993496; *Guizotia abyssinica*, EU549769; *Jacobaea vulgaris*, HQ234669; *Lactuca sativa*, AP007232) belonging to the Asteraceae family. Phylogenetic tree was constructed using MEGA6.0 (Tamura *et al.* 2013) with the parameters of neighbor - joining method and 1000 bootstrap.

2.4. Comparison of inter-species level and development of molecular markers

Sequence variations were identified by mVISTA program (<http://genome.lbl.gov/vista/mvista/submit.shtml>) and the polymorphic regions were found using MAFFT program (<http://mafft.cbrc.jp/alignment/software>). The two types of molecular markers, InDel and SNP, were developed based on the polymorphic sites in chloroplast genomes of *E. prostrata* and *E. alba*. The primers were designed using NCBI Primer-blast (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). PCR reaction for InDel markers was conducted in a total volume of 25 µL. The reaction mixtures consisted of 20 ng of template DNA, 1x PCR buffer (Vivagen, South Korea), 2.5 mM dNTP (Vivagen, South Korea), 10 pmol of each primer, and 2 units/ µL of Taq DNA polymerase. PCR conditions were as follows: 5 minutes at 95°C, following 35 cycles of 30 seconds at 95°C, 30 seconds at 58°C, 30 seconds at 72°C, and a final extension at 72°C for 5 minutes. PCR amplicons were checked on 3% agarose gel in

condition of being stained by IncloneTM Safe Gel stain. Then they were visualized under UV-trans-illuminator with a gel documentation system.

PCR reaction for SNP markers was performed in a total volume of 20 μ L. The mixtures were composed of 20 ng of template DNA, 1x PCR buffer (Vivagen, South Korea), 2.5 mM dNTP (Vivagen, South Korea), 10 pmol of each primer, 2 units/ μ L of Taq DNA polymerase, and fluorescent dye SYTO 9 (Roche Diagnostics, IN). The PCR conditions were as follows: 5 minutes at 95°C, following 35 cycles of 20 seconds at 95°C, 20 seconds at 58°C, 30 seconds at 72°C. Then, HRM analysis was conducted using LightCycler 480 (Roche Applied Science) with the conditions that were as follows: 1 minute at 95°C, 1 minute at 40°C, 5 seconds at 70°C and the temperature increased to 90°C, then cooled down to 40°C as fluorescence acquisition.

3. Results

3.1. Complete chloroplast genome sequence of *E. alba* and sequence variations between *E. prostrata* and *E. alba*

E. alba was sequenced using Illumina Miseq platform to produce 1.0 Gbp raw reads (Table 6). The assembled chloroplast genome of *E. alba* was generated by *de novo* assembly method reported in previous study (Kim *et al.*, 2015a; Kim *et al.*, 2015b).

The complete chloroplast genome sequence of *E. alba* is a circular molecule of 151,733bp long and retains quadripartite structure composed of Large Single Copy (LSC), Small Single Copy (SSC), and two Inverted Repeat (IR) regions alike other higher plant species (Lee *et al.*, 2016; Park *et al.*, 2016; Kim *et al.*, 2017). The length of LSC region is 83,300bp, and those of SSC region and IR regions are 18,283bp and 25,075bp long, respectively (Fig. 6, Table 6). A total of 114 genes were annotated in the chloroplast genome of *E. alba* including 80 protein coding genes, 30 tRNA genes, and 4 rRNA genes (Table 7). Also, GC content of the chloroplast genome of Chinese *R. chinensis* was 37.49% (data not shown).

Comparative analysis revealed 24bp differences between chloroplast genome sequences of *E. prostrata* and *E. alba*. The chloroplast genome of *E. alba* is 15bp, 24bp longer than those of *E. prostrata* in LSC region and IR regions. However, SSC region of *E. prostrata* is 63bp longer than that of *E. alba* (Table 6).

Inter-species polymorphic sites in chloroplast genomes between *E. prostrata* and *E. alba* were identified using mVISTA program (Fig. 7). Inter-species polymorphic sites are located in intergenic regions more than genic regions. A total of 58 SNPs and 29 InDels between *E. prostrata* and *E. alba* were identified. The analysis of sequence variations revealed 39 SNPs and 29

InDels were located in intergenic regions. Also, 19 InDels were identified from genic regions. These polymorphic sites were used for phylogenetic analysis of *E. alba* and development of molecular markers (Fig. 7).

Table 6. Summary of NGS data and chloroplast genome sequences of *E. prostrata* and *E. alba*.

| Collected species | Raw reads (bp) | Cp coverage (x) | Cp length (bp) | LSC length (bp) | IR length (bp) | SSC length (bp) |
|----------------------------------|-------------------|--------------------|-------------------|--------------------|-------------------|--------------------|
| <i>E. prostrata</i> ^a | 1,473,824,997 | 317.69 | 151,757 | 83,285 | 18,346 | 25,063 |
| <i>E. alba</i> | 1,027,482,009 | 89.31 | 151,733 | 83,300 | 25,075 | 18,283 |

^a conducted by Park *et al.* (2016)

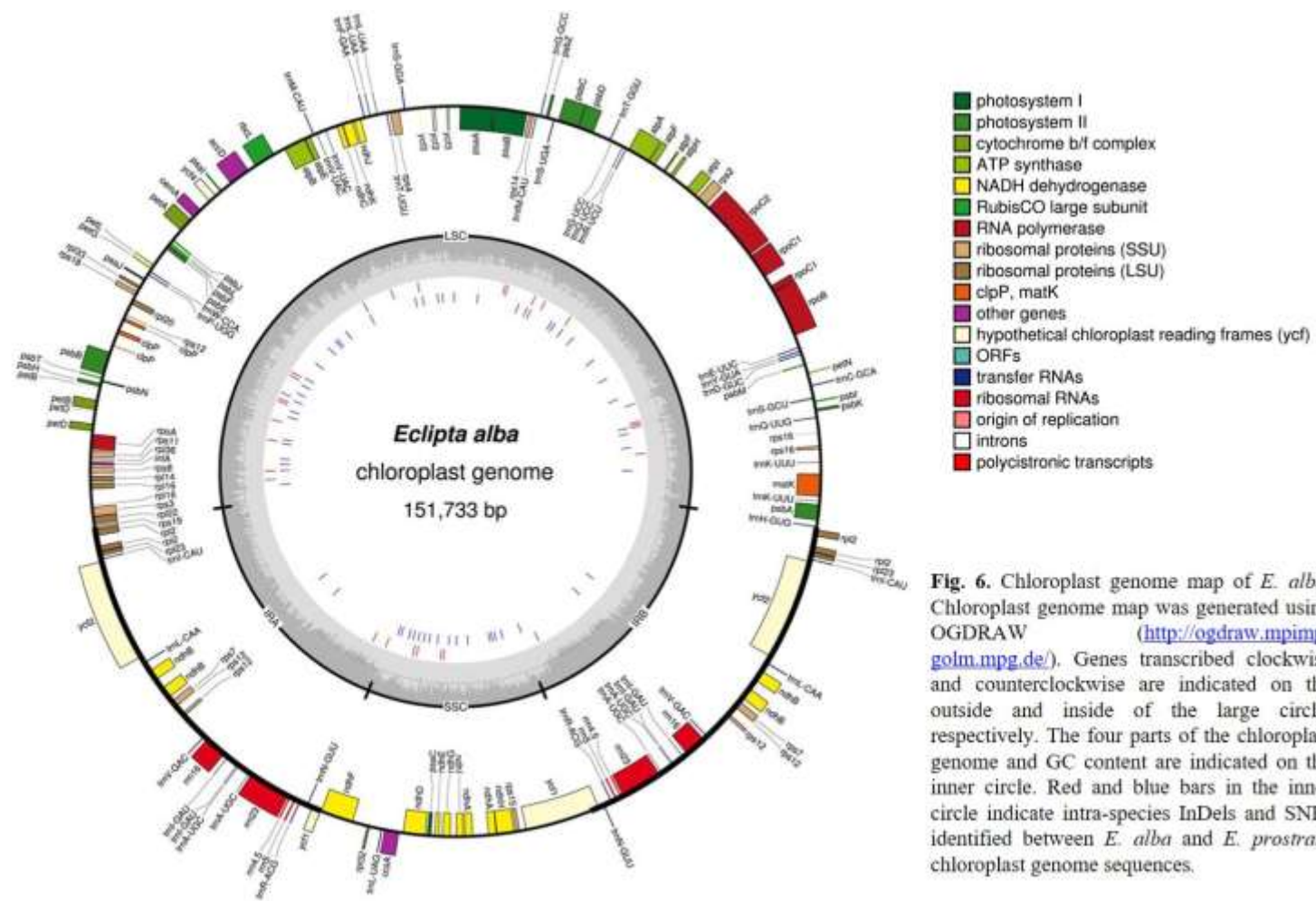


Fig. 6. Chloroplast genome map of *E. alba*. Chloroplast genome map was generated using OGDRAW (<http://ogdraw.mpimp-golm.mpg.de/>). Genes transcribed clockwise and counter-clockwise are indicated on the outside and inside of the large circle, respectively. The four parts of the chloroplast genome and GC content are indicated on the inner circle. Red and blue bars in the inner circle indicate intra-species InDels and SNPs identified between *E. alba* and *E. prostrata* chloroplast genome sequences.

Table 7. Annotated genes in chloroplast genome of *E. alba*

| Gene types | Gene production |
|---|---|
| Photosystem I | <i>psaA, B, C, I, J</i> |
| Photosystem II | <i>psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z</i> |
| Cytochrome b6/f complex | <i>petA, B, D, G, N, L</i> |
| ATP synthase | <i>atp A, B, E, F, H, I</i> |
| NADH oxidoreductase | <i>ndhA, B, C, D, E, F, G, H, I, J, K</i> |
| RubisCO large subunit | <i>rbcL</i> |
| RNA polymerase | <i>rpoA, B, C1, C2</i> |
| Ribosomal proteins (SSU) | <i>rps2, 3, 4, 7, 8, 11, 12, 14, 15, 16, 18, 19</i> |
| Ribosomal proteins (LSU) | <i>rpl2, 14, 16, 20, 22, 23, 32, 33, 36</i> |
| clpP, matK | <i>clpP, matK</i> |
| Other genes | <i>ccsA, cemA, accD, infA</i> |
| Hypothetical chloroplast reading frames | <i>ycf1, 2, 3, 4, 15</i> |
| Transfer RNAs | <i>trnA-UGC, C-GCA, D-GUC, E-UUC, F-GAA, fM-CAU, G-GCC, G-UCC, H-GUG, I-CAU, I-GAU, K-UUU, L-CAA, L-UAA, L-UAG, M-CAU, N-GUU, P-UGG, Q-UUG, R-ACG, R-UCU, S-GCU, S-GGA, S-UGA, T-GGU, T-UGU, V-GAC, V-UAC, W-CCA, Y-GUA</i> |
| Ribosomal RNAs | <i>rrn4.5, 5, 16, 23</i> |

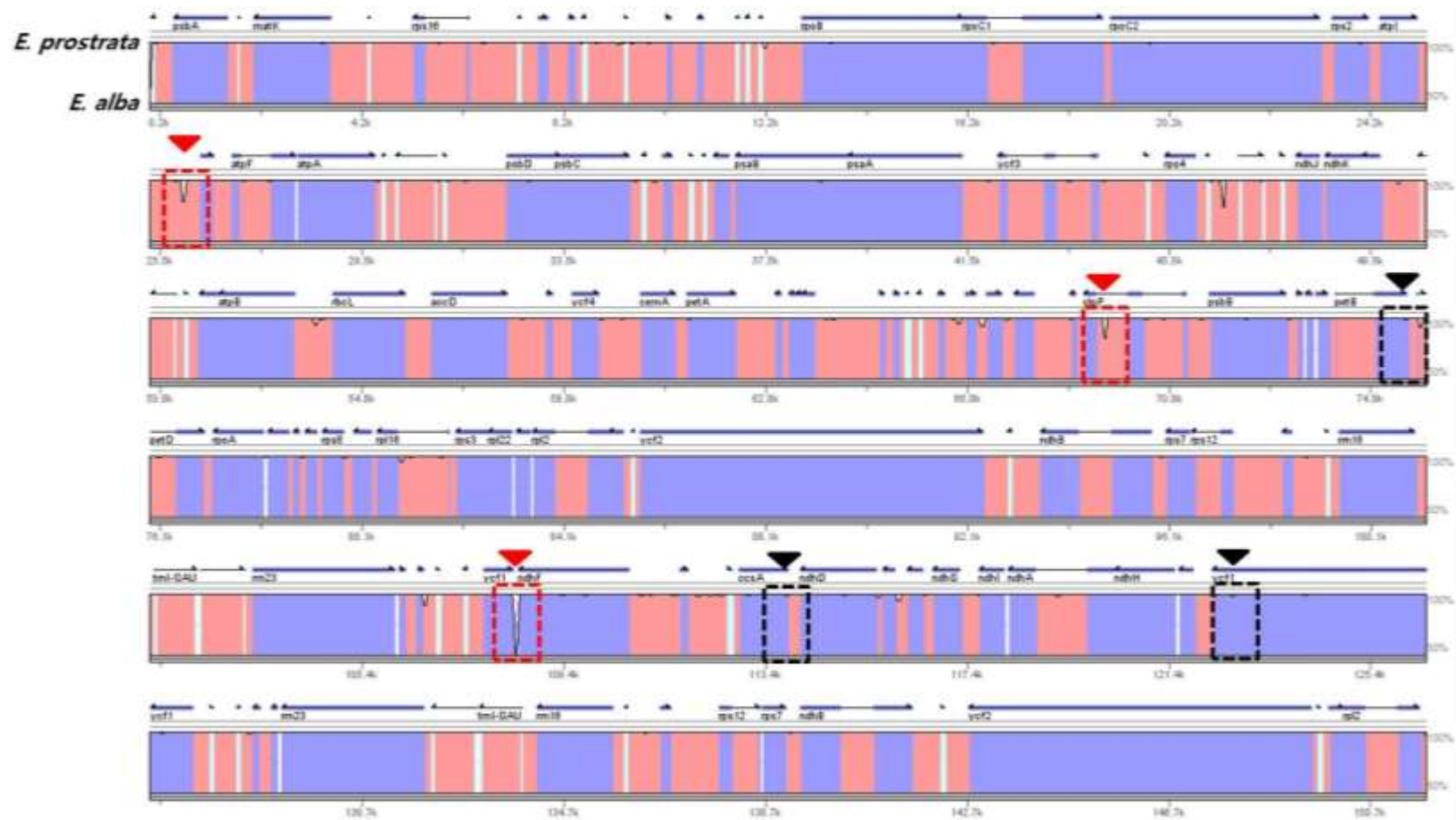


Fig. 7. Comparison of chloroplast genome sequences of *E. prostrata* and *E. alba*. Genic regions were revealed by DOGMA and BLAST search. Comparative map was generated with *E. prostrata* as a reference using mVISTA. The polymorphic regions used for molecular marker development are represented by arrowheads. Red arrowheads represent the regions used for InDel marker development. Black arrowhead shows the site used for SNP marker development

3.2. Phylogenetic analysis of *E. alba* based on the chloroplast genomes

Phylogenetic analysis was conducted with the complete chloroplast genomes of *E. prostrata* and *E. alba* & 10 complete chloroplast sequences of Asteraceae family. *E. prostrata* and *E. alba* were grouped together, as expected, with Heliantheae species (Fig. 8) and positioned with the species of Senecioneae and Cichorieae in Asteraceae family. Meanwhile, the phylogenetic tree shows that the group including *Eclipta* species exhibit close sister relationship with the species in Astereae, Gnaphalieae and Anthemideae. This analysis is in accordance with previous study of *E. prostrata* (Park *et al.*, 2016).

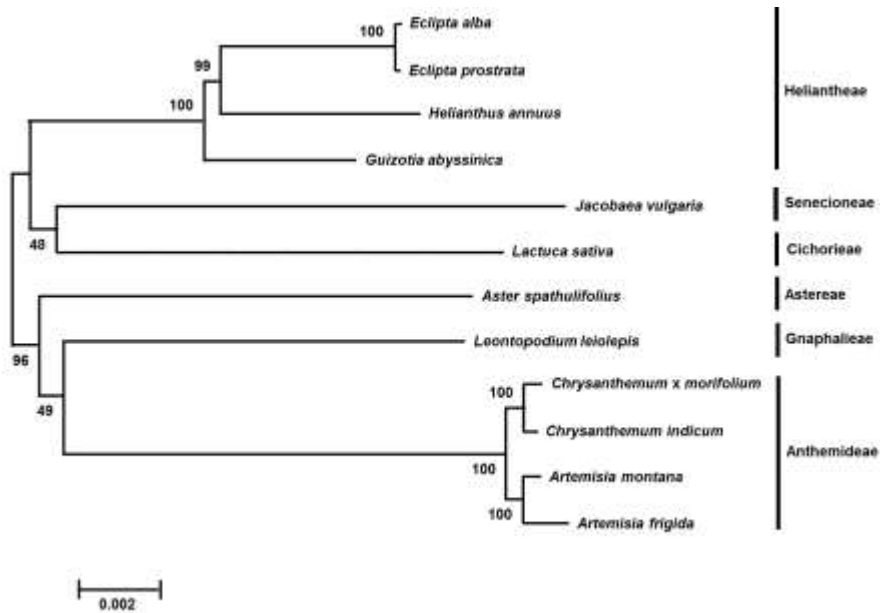


Fig. 8. Phylogenetic analysis of *E. alba*. The tree was generated with protein coding genome sequences in chloroplasts of *E. alba* and 11 species belonging to the Asteraceae family by multiple alignment using MAFFT (<http://mafft.cbrc.jp/alignment/server/index.html>) and a neighbor-joining (NJ) analysis using MEGA 6.0 (Tamura *et al.* 2013). Numbers in the nodes are bootstrap support values (>50%) from 1000 replicates. Chloroplast genome sequences used for this tree are: *Chrysanthemum x morifolium*, JQ362483; *Chrysanthemum indicum*, JN867589; *Artemisia montana*, KF887960; *Artemisia frigida*, JX293720; *Aster spathulifolius*, KF279514; *Leontopodium leiolepis*, KM267636; *Helianthus annuus*, DQ383815; *Eclipta prostrata*, KU361242; *Eclipta alba*, MF993496; *Guizotia abyssinica*, EU549769; *Jacobaea vulgaris*, HQ234669; *Lactuca sativa*, AP007232.

3.3 Molecular markers for discrimination of *E. prostrata* and *E. alba* & Application of InDel markers to 5 random specimens

The molecular markers were developed in this study for classification of at inter-species level. The information of molecular markers are described in Table 8. Amplicons produced by ep_01 showed 18-bp differences in *atpI-atpH* region between *E. prostrata* and *E. alba* as we expected (Fig. 9(A)). The InDel marker, ep_02, also successfully amplified the DNAs of *E. prostrata* and *E. alba*, and revealed 16-bp differences between those species in *clpP* region (Fig. 9(B)). Likewise, ep_03 marker demonstrated 54-bp InDel polymorphism in *ycf1-ndhF* region between *Eclipta* species as expected (Fig. 9(C)).

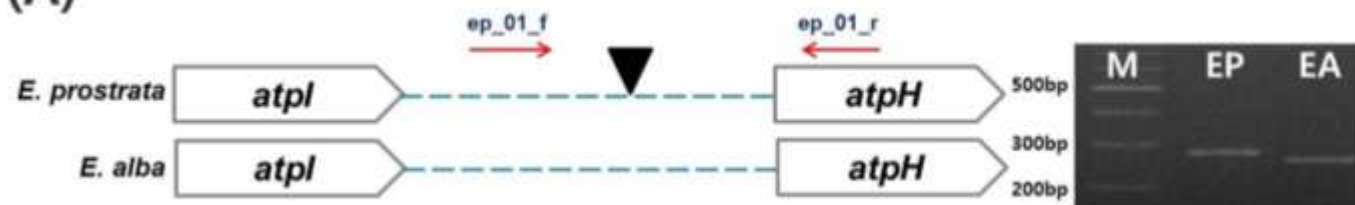
Similarly, HRM analysis applied by ep_hrm_01 showed different melting patterns of the G/A SNP in *ccsA* region of *E. prostrata* and *E. alba* (Fig. 10(A)). Also, ep_hrm_02, SNP marker, efficiently distinguished the melting patterns of T/G and G/A SNPs in *ycf1* region of chloroplast genome in *E. prostrata* and *E. alba* (Fig. 10(B)). The last SNP marker, ep_hrm_03, successfully showed different melting patterns of T/C SNP in *petB* region of *E. prostrata* and *E. alba* (Fig. 10(C)).

Furthermore, three InDel markers were applied to 5 random specimens of *Eclipta* species that are provided by Food and Drug Safety for authentication. All of the InDel markers successfully amplified the DNAs of *E. prostrata* and *E.alba*, and also the result shows both sizes of amplicons exist in the same specimen (01) (Fig. 11). The result indicates that the medicinal herbs made from *Eclipta* species are being sold with the state of mixing two species in markets.

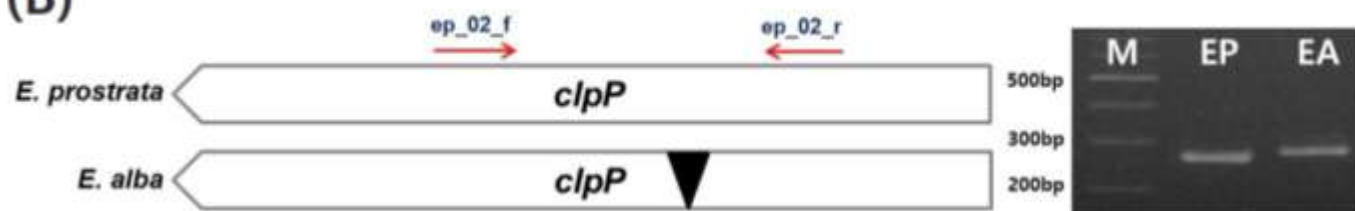
Table 8. Information of molecular markers developed in this study for classification between *E. prostrata* and *E. alba*

| Type | Marker ID | Location | | Primer Sequence | Product Size (bp) |
|-------|-----------|------------------|---|---------------------------|-------------------|
| InDel | ep_01 | <i>atpI-atpH</i> | F | TGTCAAGGGTTAGACGCATCC | 266 |
| | | | R | TGTCCCGAATCGCTCTTTGA | |
| | ep_02 | <i>clpP</i> | F | AGAACCAGCAGGTTGATGGA | 283 |
| | | | R | TTCTCCGAAAGGAAGGGTG | |
| | ep_03 | <i>ycf1-ndhF</i> | F | TCGATGCAACAGCAAGATGC | 230 |
| | | | R | AAATCATTAGGGGGTGGACG | |
| SNP | ep_hrm_01 | <i>ccsA</i> | F | GTGGCGACTCTAGGCTTTCT | 179 |
| | | | R | TCGCCGTTGAGACAAGATT | |
| | ep_hrm_02 | <i>ycf1</i> | F | TCTCTACGACGTTTAGACGATAAAA | 155 |
| | | | R | AAGCACCAAAGTAATCAAAGATACC | |
| | ep_hrm_03 | <i>petB</i> | F | TATTGCCGCTTCTTACTGCG | 146 |
| | | | R | ACTATAGTTCCTACCCCAAGTGAT | |

(A)



(B)



(C)

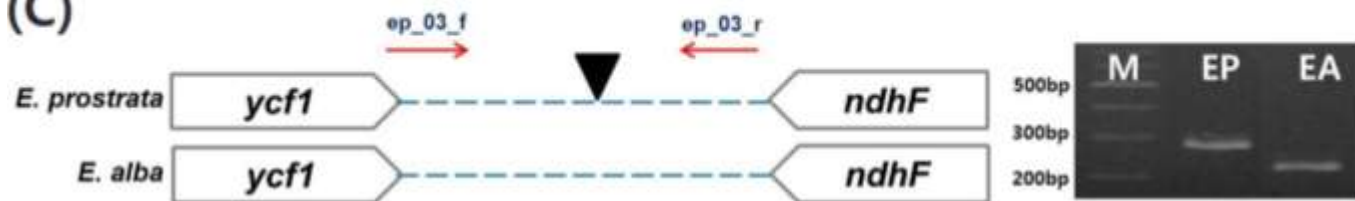
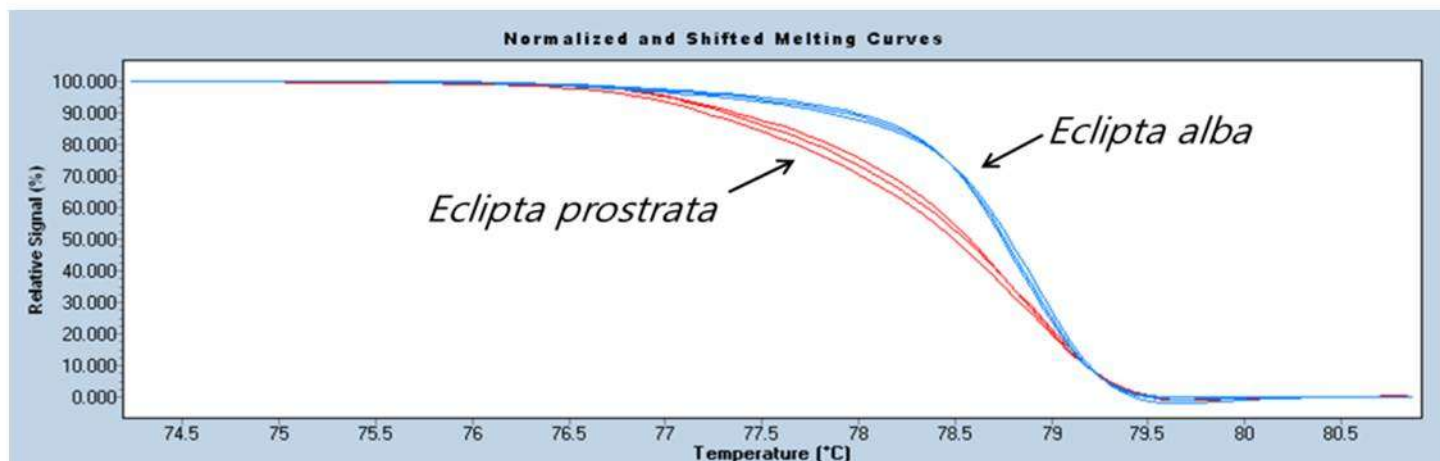
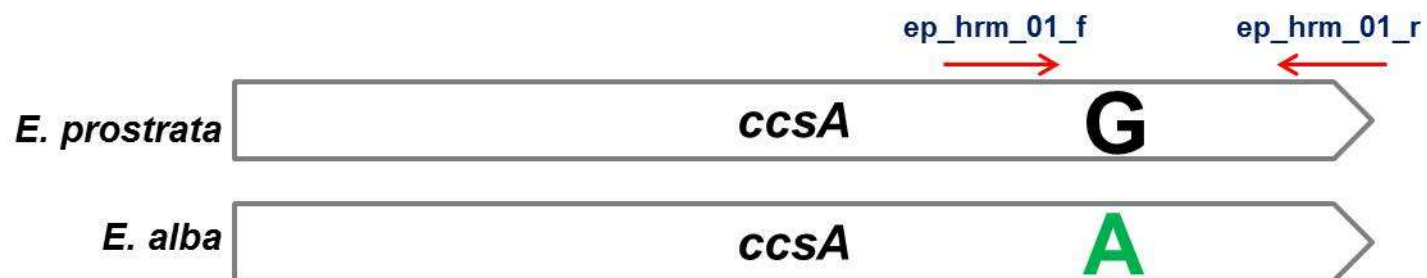
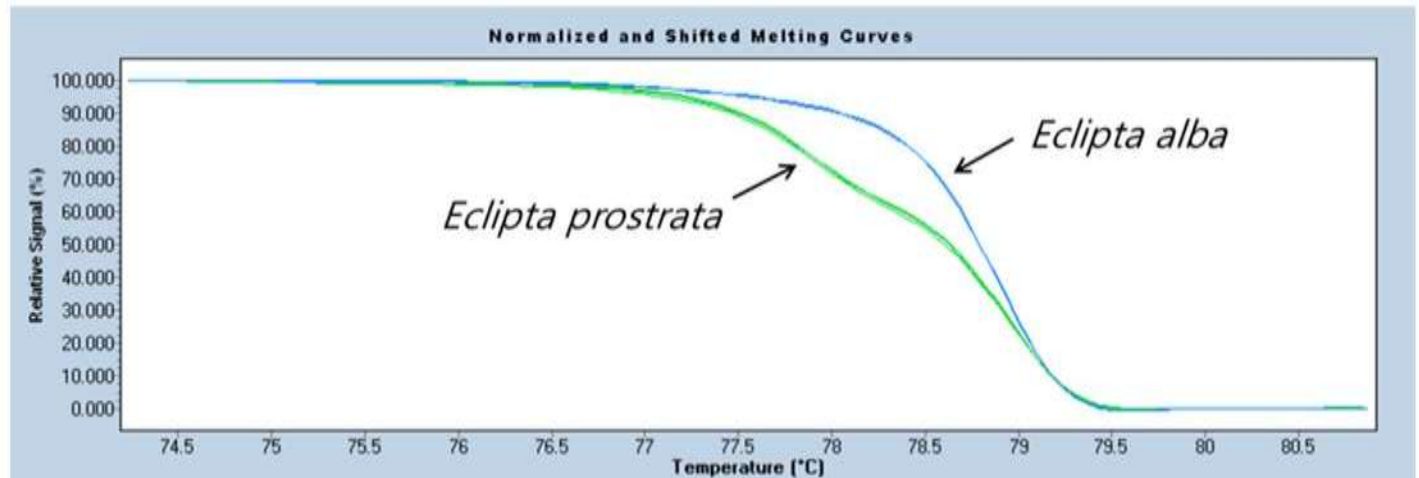
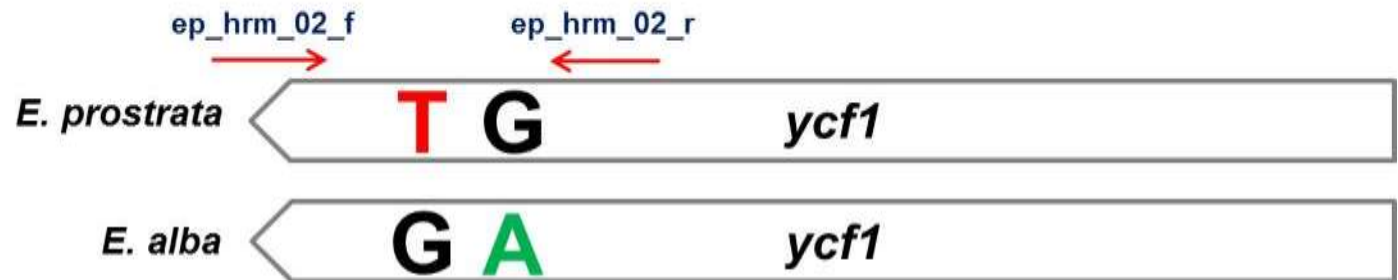


Fig. 9. InDel markers for discrimination between *E. prostrata* and *E. alba* at inter-species level. (A) The marker, ep_01 showed the amplicons that include 18-bp InDel (AATTTTCTCTACTAATCG) in *atpI-atpH* region. (B) The marker, ep_02 showed the amplicons that contain 16-bp InDel (CCTTCCATCAAATTAA) in *clpP* region. (C) The marker, ep_03 represented the amplicons that shows 54-bp InDel (AAGACTTAAAGATTAAATAAACACTTAAAAAAGACTTAAAGATTAAACTTAAAA) in *ycfI-ndhF* region. EP: *E. prostrata*, EA: *E. alba*, M: 100bp ladder.

(A)



(B)



(C)

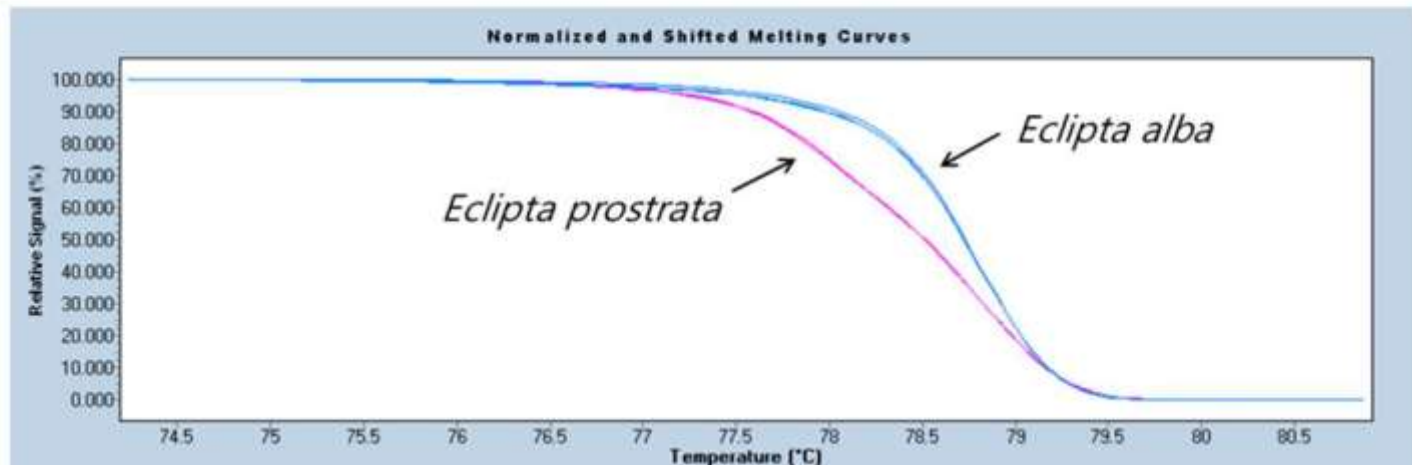


Fig. 10. SNP markers for classification between *E. prostrata* and *E. alba* at inter-species level. The markers, (A) ep_hrm_01, (B) ep_hrm_02, (C) ep_hrm_03 showed different melting patterns between *E. prostrata* and *E. alba* in *ccsA*, *ycf1*, *petB* regions, respectively.

(A)

| M | EP | EA | 01 | | | | 02 | | | | 03 | | | | 04 | 05 | M |
|---|----|----|----|---|----|----|----|---|----|----|----|---|----|----|----|----|---|
| | | | B | S | L1 | L2 | B | S | L1 | L2 | B | S | L1 | L2 | B | B | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

(B)

| M | EP | EA | 01 | | | | 02 | | | | 03 | | | | 04 | 05 | M |
|---|----|----|----|---|----|----|----|---|----|----|----|---|----|----|----|----|---|
| | | | B | S | L1 | L2 | B | S | L1 | L2 | B | S | L1 | L2 | B | B | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

(C)

| M | EP | EA | 01 | | | | 02 | | | | 03 | | | | 04 | 05 | M |
|---|----|----|----|---|----|----|----|---|----|----|----|---|----|----|----|----|---|
| | | | B | S | L1 | L2 | B | S | L1 | L2 | B | S | L1 | L2 | B | B | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |

Fig. 11. Application of InDel markers to random specimens of *Eclipta* species. The markers, (A) ep_01, (B) ep_02, (C) ep_03 were applied to *E. prostrata* and *E. alba* and the specimens in each forms of bulk, stem, leaves. EP: *E. prostrata*, EA: *E. alba*, 01-05: The names of the random specimens, B: bulk form, S: stem form, L1: leaf form, L2: another leaf form, M: 100bp ladder.

4. Discussion

4.1. Discrimination of *E. prostrata* and *E. alba* with six molecular markers

Despite different medicinal effects of *E. prostrata* from those of *E. alba*, it has been confused often with *E. alba* in the market and academic world because of their similar morphology and widely shared distribution (Baskaran and Jayabalan 2005; Dhaka and Kothari 2005). Thus, it is essential to develop a unique system to investigate *E. prostrata* and *E. alba* to protect against false trading (Neeraja and Margaret 2012).

Here, we generated the complete chloroplast genome of *E. alba* and identified 87 inter-species polymorphic sites, 58 SNPs and 29 InDels, between *E. prostrata* and *E. alba*. These polymorphic regions could be applied to phylogenetic analysis and development of molecular markers for discrimination of *E. prostrata* and *E. alba*. Phylogenetic analysis revealed that *E. prostrata* and *E. alba* are really close and grouped with Heliantheae species (Fig. 8). Overall phylogenetic result is consistent with a previous study (Park *et al.* 2016).

In this study, we report six markers that can be applied for classification of *E. prostrata* and *E. alba*, and for authentication of five random specimens of dried *Eclipta* plant tissues. Three InDel-based markers successfully distinguished *E. prostrata* from *E. alba* and three SNP markers effectively divided the melting patterns of *E. prostrata* from *E. alba* with HRM analysis (Fig. 9, 10). Overall, the results suggest that these molecular markers could contribute to ascertain the exact medicinal resources between those species.

4.2. Authentication of *Eclipta* products with three InDel-based markers

For further validation of InDel-based markers and authentication of random *Eclipta* medicinal herbs, we conducted PCR analysis with the InDel-based

markers developed in this study (Fig. 9, 11). All InDel-based markers efficiently discriminated *E. prostrata* from *E. alba* as previously proved in this research. However, sample no. 01 showed both genotypes for the markers, indicating that the products contain tissues from both species (Fig. 11).

Through this analysis, we presented an instance of admixture between *E. prostrata* and *E. alba* in the market. Both species show very similar morphology and also very low sequence level genetic diversity for complete chloroplast genomes. Relatively abundant intra-species polymorphisms are identified in *Pedicularis chamissonis*, *Panax ginseng*, *Scutellaria baicalensis* and *R. chinensis* (Fujii *et al.* 1997; Kim *et al.* 2015a; Jiang *et al.* 2017; Joh *et al.* 2017; Kim *et al.* 2017). We found very low diversity between both species with their sequence variations, suggesting an on-going genome divergence of the two species. Because of the similar morphology and the sequence level genome similarity, nevertheless, it is required to expand this research by investigating more *Eclipta* species for solid conclusion and for establishment of clear discrimination criteria.

REFERENCES

- Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc*, 1: 2320-2325.
- Baskaran P, Jayabalan N. An efficient micropropagation system for *Eclipta alba*— a valuable medicinal herb. (2005) *In Vitro Cellular and Developmental Biology-Plant*, 41(4): 532-539.
- Cheng Y, de Vicente MC, Meng H, Guo W, Tao N, Deng X. (2005) A set of primers for analyzing chloroplast DNA diversity in *Citrus* and related genera. *Tree Physiol*, 25: 661-672.
- Chinese Pharmacopoeia Commission (2010) *Pharmacopoeia of the People's Republic of China 2010*. Chemical Industry Press.
- Chokotia LS, Vashistha P, Sironiya R, Matoli H. (2013) Pharmacological Activities of *Eclipta alba* (L.). *International Journal of Research and Development in Pharmacy and Life Sciences*, 2(4): 499-502.
- Dalal S, Kataria SK, Sastry KV, Rana SVS. (2010) Phytochemical screening of methanolic extract and antibacterial activity of active principles of hepatoprotective herb, *Eclipta alba*. *Ethnobotanical leaflets*, 14: 248-258.
- Datta K, Singh AT, Mukherjee A, Bhat B, Ramesh B, Burman AC. (2009) *Eclipta alba* extract with potential for hair growth promoting activity. *Journal of Ethnopharmacology*, 124(3): 450-456.
- Dhaka N, Kothari SL. (2005) Micropropagation of *Eclipta alba* (L.) Hassk—an important medicinal plant. *In Vitro Cellular and Developmental Biology-Plant*. 41(5): 658-661.
- Djakpo O, Yao W (2010) *Rhus chinensis* and *Galla Chinensis*--folklore to modern evidence: review. *Phytother Res*, 24:1739-1747.

- Dong W, Liu J, Yu J, Wang L, Zhou S. (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One, 7: e35071.
- Fujii N, Ueda K, Watano Y & Shimizu T. (1995) Intraspecific sequence variation in chloroplast DNA of *Primula cuneifolia* Ledeb (Primulaceae). J. Phytogeogr. Taxon, 43(1-2): 15-24.
- Fujii N, Ueda K, Watano Y & Shimizu T. (1997) Intraspecific sequence variation of chloroplast DNA in *Pedicularis chamissonis* Steven (*Scrophulariaceae*) and geographic structuring of the Japanese “Alpine” plants. Journal of Plant Research, 110(2): 195-207.
- Grisi MCM, Blair MW, Gepts P, Brondani C, Pereira PAA & Brondani RPV. (2007) Genetic mapping of a new set of microsatellite markers in a reference common bean (*Phaseolus vulgaris*) population BAT93 x Jalo EEP558. Genetics and Molecular Research, 6(3): 691-706.
- Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, Bank MVD, *et al.* (2009) A DNA barcode for land plants. PNAS, 106(31): 12794-12797.
- Jang WJ, Kim KY, Kim KH, Lee SC, Park HS, Lee JK, *et al.* (2015) The complete chloroplast genome sequence of *Cynanchum auriculatum* Royle ex Wight (*Apocynaceae*). Mitochondrial DNA Part A, 6: 4549-4550.
- Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC *et al.* (2005) Methods in Enzymology, 395:348-384.
- Jiang D, Zhao Z, Zhang T, Zhong W, Liu C, Yuan Q, *et al.* (2017) The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. Genes, 8: 227.
- Joh HJ, Kim NH, Jayakodi M, Jang WJ, Park JY, Kim YC, *et al.* (2017) Authentication of Golden-Berry *P. ginseng* Cultivar ‘Gumpoong’ from a Landrace ‘Hwangsook’ Based on Pooling Method Using Chloroplast-Derived Markers. Plant Breeding Biotechnology, 5(1): 16-24.

- Khairallah ARAT & Salama B. (2009) Genetic variation among sumac (*Rhus Coriaria* L.) samples collected from three locations in Jordan as revealed by AFLP markers. *Advances in Environmental Biology*, 3(1): 107-112.
- Kim I, Park JY, Lee YS, Joh HJ, Kang SJ, Jayakodi M, *et al.* (2017) The complete chloroplast genome sequence and intra-species diversity of *Rhus chinensis*. *Plant Breed. Biotech*, 5: 243-251.
- Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim NH, *et al.* (2015a) Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PLoS one*, 10(6): e0117159.
- Kim K, Lee SC, Lee J, Yu Y, Yang K, Choi BS, *et al.* (2015b) Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci Rep*, 5:15655.
- Kim KH, Kim YS, Kim MR, Lee HY, Lee KH, Kim JH, *et al.* (2015c) Development of primer sets for the detection of *Polygonum multiflorum*, *Cynanchum wilfordii* and *C. auriculatum*. *Journal of Food Hygiene and Safety*, 30: 289-294.
- Lee WK, Kim MJ & Heo K. (2004) Phylogeny of Korean *Rhus* spp. based on ITS and *rbcL* sequences. *Korean Journal of Medicinal Crop Science*, 12(1): 60-66.
- Lee YS, Kim IS, Kim JK, Park JY, Joh HJ, Park HS, *et al.* (2016) The complete chloroplast genome sequence of *Rhus chinensis* Mill (Anacardiaceae). *Mitochondrial DNA Part B*, 1:696-697.
- Lim CY, Oh CH (2015) Growth and spatial distribution characteristics of *Rhus javanica* populations sowed on cut-slopes. *J. Korean Env Res Tech*, 18: 39-49.
- Ling KS, Abdullah SNA, Ling HC, Amiruddin MD. (2016) Molecular cloning, gene expression profiling and *in silico* sequence analysis of vitamin E biosynthetic genes from the oil palm. *Plant Gene*, 5:100-108.

- Liu QM, Zhao HY, Zhong XK, Jiang JG. (2012) *Eclipta prostrata* L. phytochemicals: isolation, structure elucidation, and their antitumor activity. *Food Chem. Toxicol*, 50: 4016-4022.
- Ma-Ma K, Nyunt N, Tin KM. (1978) The protective effect of *Eclipta alba* on carbon tetrachloride-induced acute liver damage. *Toxicology and applied pharmacology*, 45(3): 723-728.
- Ma WL, Wu M, Wu Y, Ren ZM, Zhong Y. (2013) Cloning and characterisation of a phenylalanine ammonia-lyase gene from *Rhus chinensis*. *Plant Cell Rep*, 32:1179–1190.
- Miller AJ, Young DA, Wen J. (2001) Phylogeny and biogeography of *Rhus* (Anacardiaceae) based on ITS sequences data. *Int J Plant Sci*, 162: 1401–1407.
- Min T, Barfod A (2008) Anacardiaceae. In: Wu ZY, Raven PH, Hong DY (eds) *Flora of China*. Vol. 11. Beijing, China and St. Louis, Missouri, (MO), USA: Science Press and Missouri Botanical Garden Press. http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=10038.
- Mithun NM, S. Shashidhara, R. Vivek Kumar. (2011) *Eclipta alba* (L.) A review on its phytochemical and pharmacological profile. *Pharmacologyonline*, 1: 345-357.
- Mora-Ortiz M, Swain MT, Vickers MJ, Hegarty MJ, Kelly R, Smith LMJ *et al.* (2016) *De-novo* transcriptome assembly for gene identification, analysis, annotation, and molecular marker discovery in *Onobrychis viciifolia*. *BMC Genomics*, 17:756.
- Muruganantham S, Anbalagan S, Ramamurthy N. (2009) FT-IR and SEM-EDS comparative analysis of medicinal plants, *Eclipta alba* Hassk and *Eclipta prostrata* Linn. *Romanian J. Biophys*, 19(4): 285-294.
- Nam SJ, Yeo WJ, Choi JH, Kim NC. (2004) Development of revegetation method using forest topsoils for ecological restoration of the slopes (I). *J of Korean Env Res & Reveg Tech*, 7:110-119.

- Neeraja PV, Margaret E. (2012) *Eclipta alba* (L.) Hassk: a valuable medicinal herb. International Journal of Current Pharmaceutical Review and Research, 2: 188-197.
- Nielsen LR. (2004) Molecular differentiation within and among island populations of the endemic plant *Scalesia affinis* (Asteraceae) from the Galápagos Islands. Heredity, 93(5): 434-442.
- Parducci L, Szmidt AE, Madaghiele A, Anzidei M, Vendramin GG. (2001) Genetic variation at chloroplast microsatellites (cpSSRs) in *Abies nebrodensis* (Lojac.) Mattei and three neighboring *Abies* species. Theoretical and Applied Genetics, 102(5): 733-740.
- Park JY, Lee YS, Kim JK, Lee HO, Park HS, Lee SC, *et al.* 2016. The complete chloroplast genome of *Eclipta prostrata* L.(Asteraceae). Mitochondrial DNA Part B, 1(1): 414-415.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, *et al.* (2000) Artemis: sequence visualization and annotation. Bioinformatics, 16: 944-945.
- Sewell MM, Parks CR & Chase MW. (1996) Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L.(Magnoliaceae). Evolution, 1147-1154.
- Soltis D, Soltis P, Kuzoff R & Tucker T. (1992) Geographic structuring of chloroplast DNA genotypes in *Tiarella trifoliata* (Saxifragaceae). Plant Systematics and Evolution, 181(3): 203-216.
- Soltis DE, Mayer MS, Soltis PS & Edgerton M. (1991) Chloroplast-DNA variation in *Tellima grandiflora* (Saxifragaceae). American Journal of Botany, 1379-1390.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol., 30: 2725-2729.

- Yang JB, Li DZ, Li HT. (2014) Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour*, 14:1024-1031.
- Yi T, Miller AJ, Wen J. (2004) Phylogenetic and biogeographic diversification of *Rhus* (Anacardiaceae) in the Northern Hemisphere. *Mol Phylogenet Evol.*, 33:861-879.
- Yi T, Miller AJ, & Wen J. (2007) Phylogeny of *Rhus* (Anacardiaceae) based on sequences of nuclear *Nia-i3* intron and chloroplast *trnC-trnD*. *Systematic Botany*, 32(2): 379-391.
- Young DA. (1978) Re-evaluation of the section of *Rhus* L. subgenus *Lobadium* (Raf.) T. and G. (Anacardiaceae). *Brittonia*, 30:411–415.
- Zhou W, Tang Z, Hou J, Hu N, Yin T. (2015) Genetic Map Construction and Detection of Genetic Loci Underlying Segregation Distortion in an Intraspecific Cross of *Populus deltoides*. *PLoS ONE*, 10(5): e0126077

ABSTRACT IN KOREAN

불나무는 윣나무과에 속한 관목으로 아시아 전역에 분포한다. 이 식물은 약용식물과 생태 복원 종으로 사용된다. 또한, 한련초는 국화과에 속하는 약용식물이다. 한련초는 가는잎한련초와 비슷한 외형적 특징과 서식지로 인해 자주 혼동되어 사용된다. 이 식물들을 종내, 중간 수준에서 구분하기 위하여 본 실험에서 불나무의 중국수집종 엽록체를 분석하였다. 불나무 중국수집종의 엽록체 서열 길이는 149,094bp 이고, Large Single Copy (97,246bp), Small Single Copy (18,644bp), 그리고 Inverted Repeat (16,602bp)로 이루어져있다. 이 불나무 엽록체에는 77 개의 protein coding genes, 30 개의 tRNA genes, 그리고 4 개의 rRNA genes 를 포함하여 총 111 개의 유전자가 존재하는 것으로 밝혀졌다. 불나무와 무환자나무목에 속하는 11 개 종의 완성된 엽록체 서열을 바탕으로 한 계통학적 분석은 불나무와 무환자나무목에 속하는 종들간의 관계를 입증하였다. 불나무의 비교 분석을 통해 종내수준에서 170 개의 SNP 와 85 개의 InDel 이 확인되었다. 불나무의 중국수집종과 한국수집종의 서열간 변이들을 기반으로 종내수준에서 불나무의 유전적 다양성을 확인할 수 있는 분자 마커 세 개를 개발하였다. 이와 같은 방법으로, 본 실험에서

가는잎한련초의 엽록체도 분석하였다. 가는잎한련초의 엽록체 서열 길이는 151,733bp 이고 붉나무와 같은 엽록체의 구조를 갖추고 있다. Large Single Copy, Small Single Copy, 그리고 Inverted Repeat 지역은 각각 83,300bp, 18,283bp, 그리고 25,075bp 였다. 가는잎한련초의 엽록체에는 80 개의 protein coding genes 과 30 개의 tRNA genes, 그리고 4 개의 rRNA genes 들이 존재하는 것으로 밝혀졌다. 가는잎한련초의 엽록체와 국화과에 속하는 12 개의 완성된 엽록체 서열을 대상으로 실시한 계통학적 분석은 가는잎한련초와 국화과에 속하는 식물들의 관계를 확인시킬 수 있었다. 한련초와 가는잎한련초 간의 비교분석을 통해 중간 수준에서 58 개의 InDel 과 29 개의 SNP 가 존재하는 것이 확인되었다. 이러한 다형성을 보이는 지역을 기반으로 한련초와 가는잎한련초를 효과적으로 구분할 수 있는 분자마커 6 개를 개발하였다. 본 연구에서 완성한 약용식물들의 엽록체 서열들과 종 구별 및 확인을 위해 제작된 분자마커들은 종내 및 중간 수준에서 식물의 올바른 개체 선정에 기여할 수 있을 뿐만 아니라 향후 유전적 연구에 중요한 정보를 제공할 수 있을 것이다.